USE OF ECMWF PRODUCTS AND PERFORMANCE OF THE FORECASTING SYSTEM

F. Lalaurette
European Centre for Medium-range Weather Forecasts
Reading, England

Summary: The rather unique position of ECMWF in the NWP world as a centre only devoted to medium-range forecasting made it a good place where meteorologists could exchange experience and develop methods suitable for these forecast ranges. Before the EPS was developed, it was an important component of the medium range forecast to identify predictable scales. An important message conveyed to the users was then that small scale phenomena were unpredictable in the medium range more as a result of intrinsic unpredictability than by model inconsistency that could be solved by improvements of the physics or numerics. Now the EPS provides a direct, flow dependent assessment of the predictability. The first part of this paper describes how new methods were developed in close link with Member States forecasters to make use of these new products. The question of how we can validate ensemble forecasts, both in deterministic and in probabilistic mode, is also addressed.

The improvement in the representation of physical processes in the model together with the increased resolution now makes it possible to measure the model performance by assessing directly its ability to forecast parameters related to the weather. The second part of the paper shows how the model performs in the short range when forecasting temperature, wind and precipitation, and how much skill can be found also in the medium range when the EPS is used.

1 INTRODUCTION

Focusing on the medium range forecasting gave ECMWF a unique opportunity over the years to develop post-processing and interpretation methods tailored for applications in this range. The role of the Centre as a focal point for discussions and exchange of experience was played during the bi-annual Workshops on Meteorological Operational Systems, Training Courses and Expert Meetings held with participation from European Member States and other leading centres in the world. Thinking "Medium-Range" has made it possible to develop specific, supra-synoptic interpretation methods of interpretation that ultimately have culminated with the development of the EPS. The development of products for the EPS has been another fruitful opportunity of exchanges with the forecasters and other users in the Member States. Following a period when only pioneers dared look at these new forecasts, time is now ripe to draw the first conclusions of a widely use of the EPS and to re-design some of the products offered to our users.

Another specific aspect of ECMWF is that verification is under the responsibility of the same group that works on the design of new products and interpretation methods. It is also a strong belief that the interpretation should always be carefully constrained by our quantitative knowledge of model errors. In the case of the EPS and more generally of probabilistic forecast products, it can even be shown that the value that can be extracted from the system is closely associated with the knowledge of error characteristics, as was demonstrated by *Richardson* (2000) for example.

2 THINKING "MEDIUM RANGE"

2.1 **Pre-EPS ages**

The need for a "supra-synoptic" interpretation of the medium range forecast slowly emerged from the accumulation of experience gathered by the first users of medium range numerical products. The first concerns were expressed by the forecasters: they quickly complained that although the Day 10 forecast was looking as realistic as a Day 1, errors were much more likely to occur in the former than in the latter. The reason unfortunately could not only be traced to model systematic errors that would accumulate in time and that the forecaster could correct. The model results were found to be quite inconsistent, small changes brought to the analysis every day bringing large differences in the medium range forecast. The development of error propagation visualisation techniques confirmed this strikingly. A more objective, quantitative measure of the relation between error and forecast range can be provided by a spectral analysis of the errors. An example is shown in Figure 1, where it can be seen that on average, errors in the medium range affect a wide range of scales - at Day 6, the signal over noise ratio is much less than one for wave numbers 20 and higher. More on this can be found in *Persson,* 2000.

The first obvious - if not uncontroversial -way of turning this knowledge of error scale/forecast range dependency into operational practice was to promote the use of filtered fields to isolate the robust part of the forecast and only give fine resolution patterns an indicative value. An example of filtered field is provided in Figure 2. Such maps however could hardly meet most users needs. The reluctance of forecasters to make use of them partly came from the incorrect assumption that such maps should be used as non-filtered maps when deriving weather elements. If such a map as found in Figure 2 was a non-filtered one, it is likely that no weather event of importance (strong winds, heavy precipitation) would ever be forecasted. However quite a different interpretation is to see it as an *average of many equally likely scenarios*, most of them carrying significant weather elements but with varying location and intensity, such that they do not show up in the average. A first explicit need for an *ensemble* prediction system where these varying weather elements would be explicitly represented therefore emerged. Another reason for the need of a more refined evaluation of the error scale/forecast range relationship was that this relation was found to be strongly dependent on the flow itself.

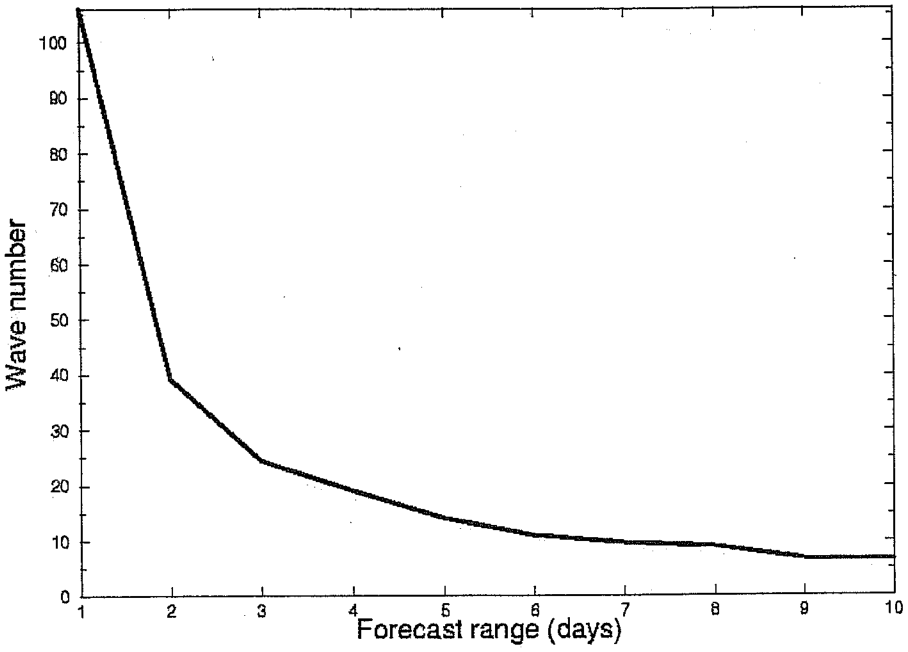
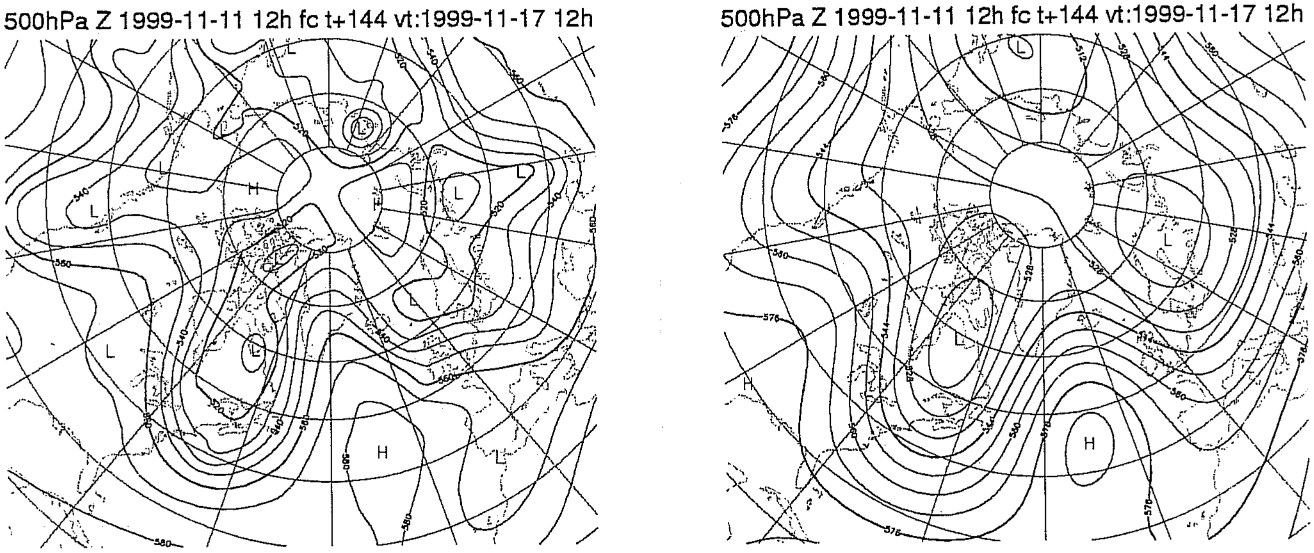


Figure 1: Wavenumber for which the energy of the forecast error at a given forecast range exceeds the energy in the analysis (Z500, Dec. 1993)


500hPa Z 1999-11-11 12h fc t+144 vt:1999-11-17 12h

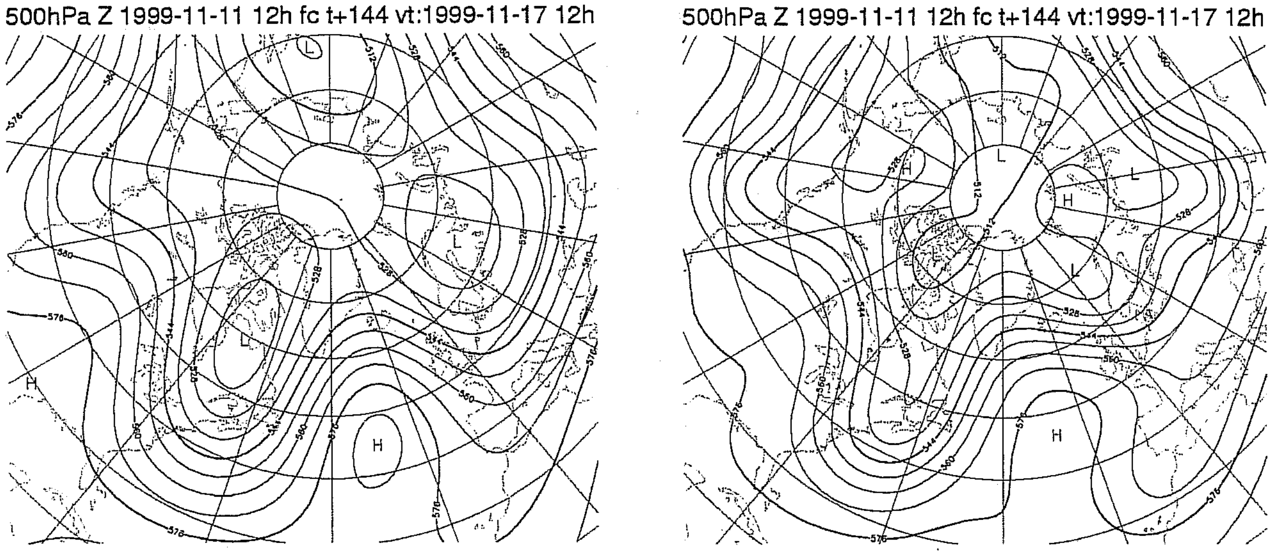
500hPa Z 1999-11-11 12h fc t+144 vt:1999-11-17 12h

Figure 2: Effect of filtering small scales (wavenumber 10 and higher) on the Day6 Z500 operational forecast Base 11 November 1999; the left panel is unfiltered, the right one is filtered

2.2 Early EPS ages

Considerations of the scale/range predictability relationships provided a natural framework in the early stages of the EPS. Indeed the use of the EPS Ensemble Mean was a natural way to move from a "blind" filter based on the average error characteristics to a more clever, flow dependent filter that would react in a different way whether the flow is predictable or not. An example of the benefit that can be expected from this approach is provided in Figure 3 where both types of filters are applied. Although the main feature remain the same, stronger gradients and even some small scale systems (e.g. the low North West of Greenland or off the Norwegian coast) are kept in the ensemble mean that were filtered out in the T10 filter. The impact of filtering through the Ensemble Mean can be found in scores, and can be furthermore demonstrated to be better than any "blind" filter (*Bouteloup*, 1997).

Reducing the EPS to the mere production of an ensemble mean would however be largely underestimating the amount of useful information that can be derived for this set of highly non-linearly related products. The reason why the emphasis was kept pretty long on such products is because most forecasters had derived their expertise in the medium range from the careful examination of the large scale flow. The 500-geopotential distributions being usually smooth and Gaussian, the ensemble mean and standard deviation were proven to give a pretty fair description of the ensemble. Indeed clustering algorithms proved to provide only poorly defined alternatives, the existence of truly multimodal distributions in the EPS being far from frequent.


500hPa Z 1999-11-11 12h fc t+144 vt:1999-11-17 12h

500hPa Z 1999-11-11 12h fc t+144 vt:1999-11-17 12h

Figure 3: Comparison between the T10 filter (left) and the Ensemble Mean (right); the forecast is the same as in Figure 2)

Things however turn out to be rather different when one tries to make use of more non-linear and skewed distributions such as those for precipitation. The dissemination by fax of "plume" products indeed proved very popular among the users, as they offered a clear indication of how much likely the occurrence of rain would be in the forecast. Ideally the dissemination of probability maps should also have been a popular way of carrying the EPS information. However the forecasters clearly expressed the need of more guidance on how to translate this information into making decisions when facing the users.

2.3 The EPS mature age

In order to build up on the success of the plumes among the users and to combine it in a format close to another popular product (Meteograms), the EPSgrams have been developed (Figure 4). Unlike for the plumes, any dynamical history is lost in the EPSgrams to only retain the statistical aspects of the distribution at any time in the forecast.

A further example of such a sampling of the EPS distribution in a statistical rather than dynamical way can be found in Figure 5. On these maps the probability is a parameter when for usual probability maps the meteorological value is a parameter and the probability is the variable. An advantage of these maps is that there is no need of an a-priori knowledge of the climate distributions in order to decide of reasonable meteorological values for the event on which to compute the probabilities. If the domain is covering areas with very different climate properties (as is the case for parameters bearing strong land/sea contrasts), this is likely to be a good way to post-process the EPS information.

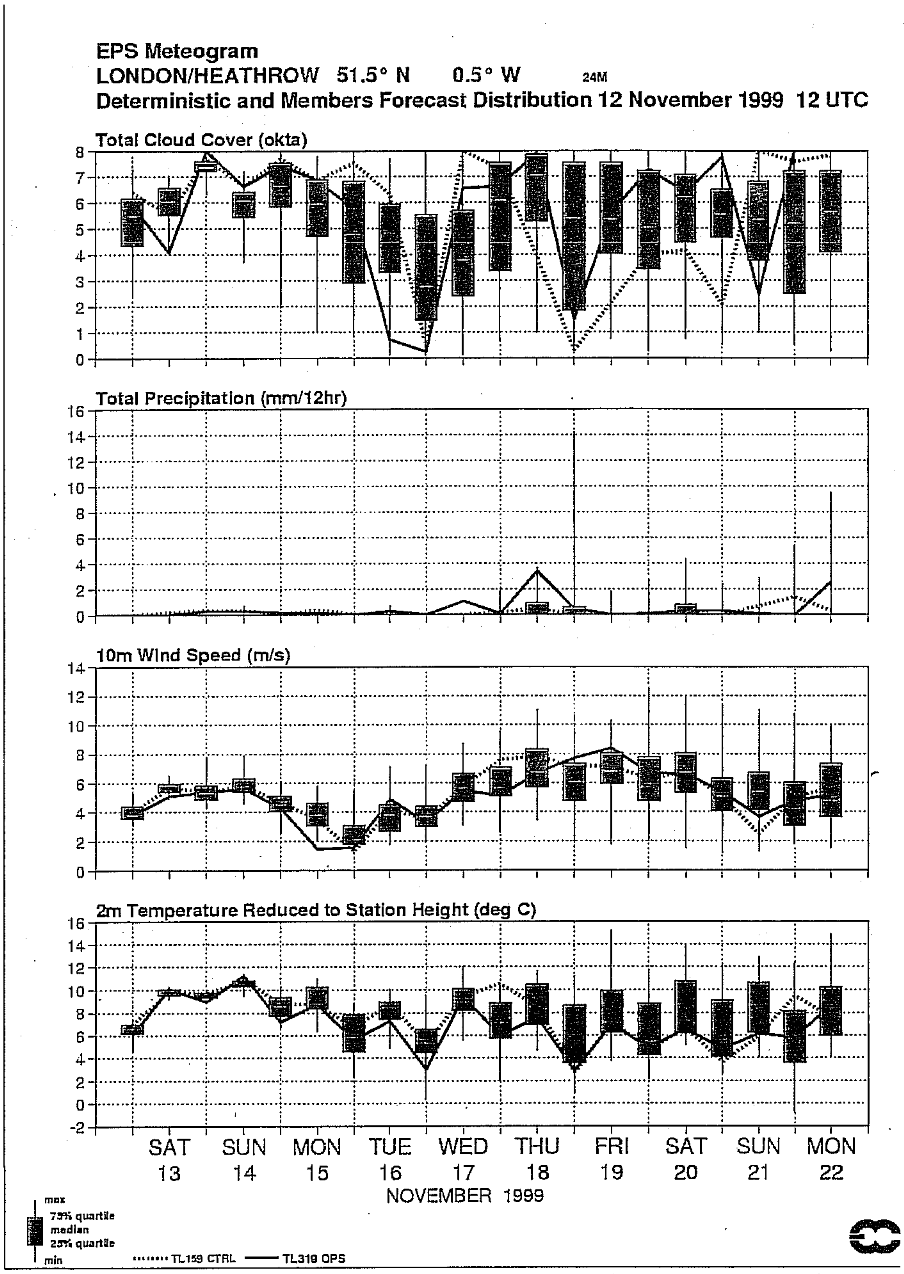


Figure 4: An example of EPSgram (London Heathrow, forecast base is 12 November 1999)

3 VALIDATION OR VERIFICATION?

3.1 Introduction

It is sometimes confusing to talk about forecast verification as this is an area where the requirements can vary a lot depending whether you are facing scientists looking for an assessment of the model performance or users looking for the value of the forecast products. In that respect, one can introduce a distinction between *validation* that is checking whether the physical processes in the numerical model fit with what can be observed and *verification* which is the measure by the end user that the forecast he was given fits with its own independent observation. In the former case, observations will be processed to be presented to the validation in a format matching as closely as possible the model variables. In the latter, maximum efforts will be spent to post-process the model output to put it in a format as close as possible to the user's observations (Figure 6).

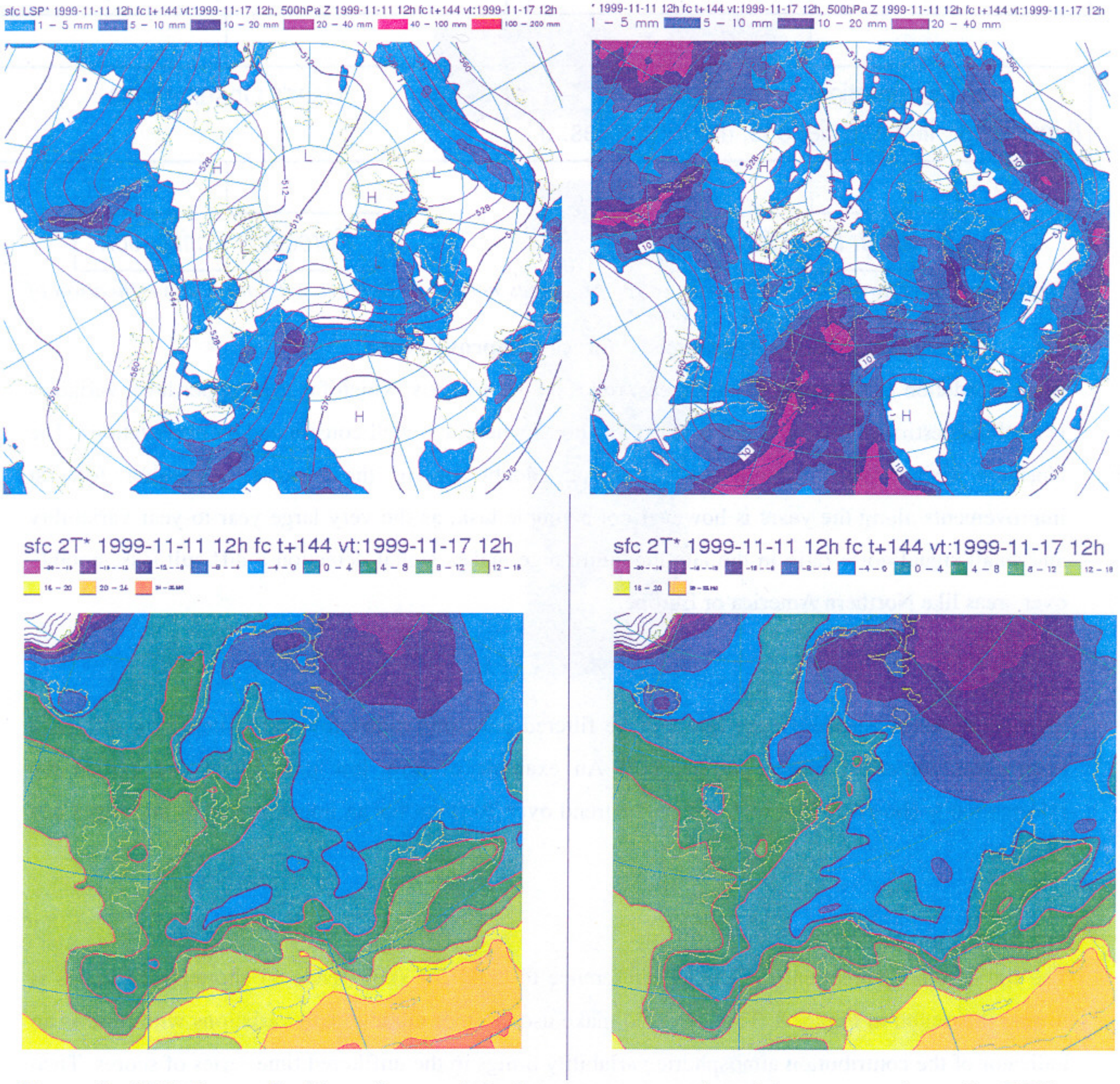


Figure 5: EPS Percentile Maps for precipitation (upper pannels, with contours from the Z500 Ensemble Mean added) and 2m-temperature (lower pannels); Left pannels show the median; the right ones show the last decile for precipitation, the first (cold) one for T2m;

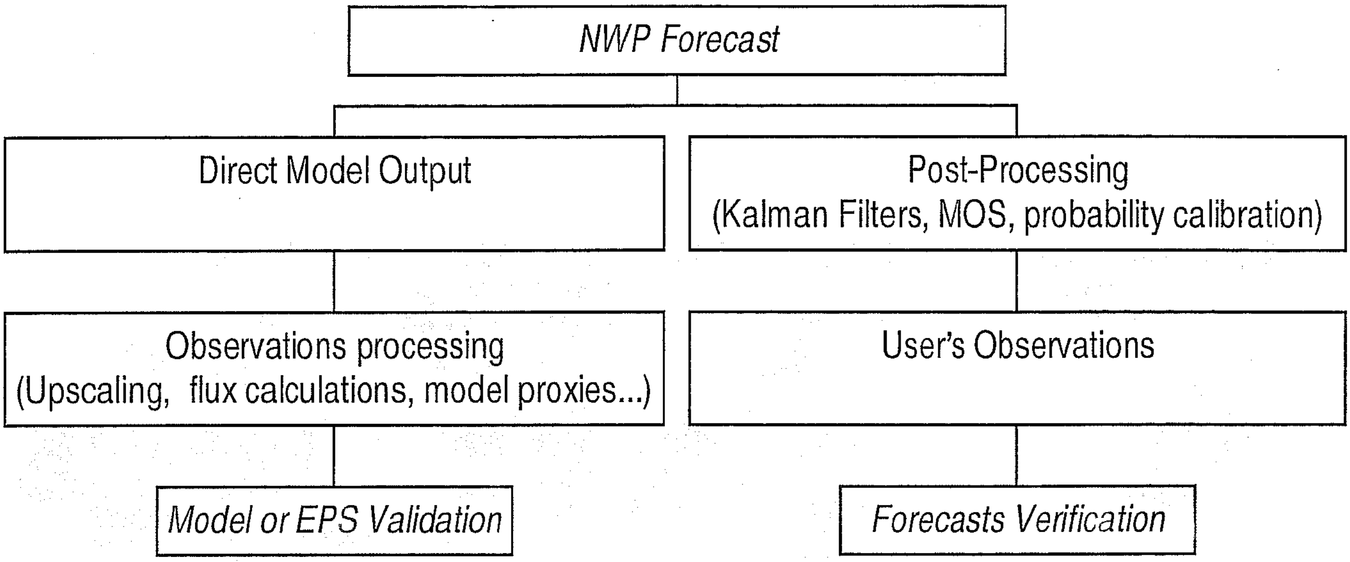


Figure 6: Chart flow showing the respective processes needed to achieve validation or verification of the numerical forecast.

3.2 Validation of recent deterministic forecast improvements

Basic validation of the operational forecast relies on the analysis, which for most upper level variables is a reliable estimate provided that it is used in areas where it is well constrained by observations. The Northern Extratropical troposphere is without doubt among those areas. Assessing forecast improvements along the years is however not a simple task, as the very large year-to-year variability and small signal/noise ratio in the late medium range makes signal detection a difficult task, notably over areas like Northern America or Europe.

This atmospheric variability can however be filtered out using either low-pass filters or a reference from other numerical model performance. An example of how low pass filters can extract the information is shown in Figure 7, where the trend over 20 years for an improvement at all ranges can be seen.

The use of other numerical models as a reference to highlight skill trends on shorter time scales is further illustrated in Figure 8. The idea is to make use of other models' skill variations with time as an indicator of the contribution atmospheric variability brings in the unfiltered time series of scores. Then the skill variability can be separated into the part that is common to most other models, and that part that is specific to the model of interest (ECMWF/IFS in our case). This kind of processing clearly helps with more objective assessment of periods of bad performance when further investigation is requested - although periods of outstanding performance also show up remarkably! Even by filtering in this way the atmospheric component in the time variations in skill, the correlation between major model changes and improvement in performance is only rarely found - the ultimate validation of model changes therefore remains the parallel testing of operational and pre-operational model versions.

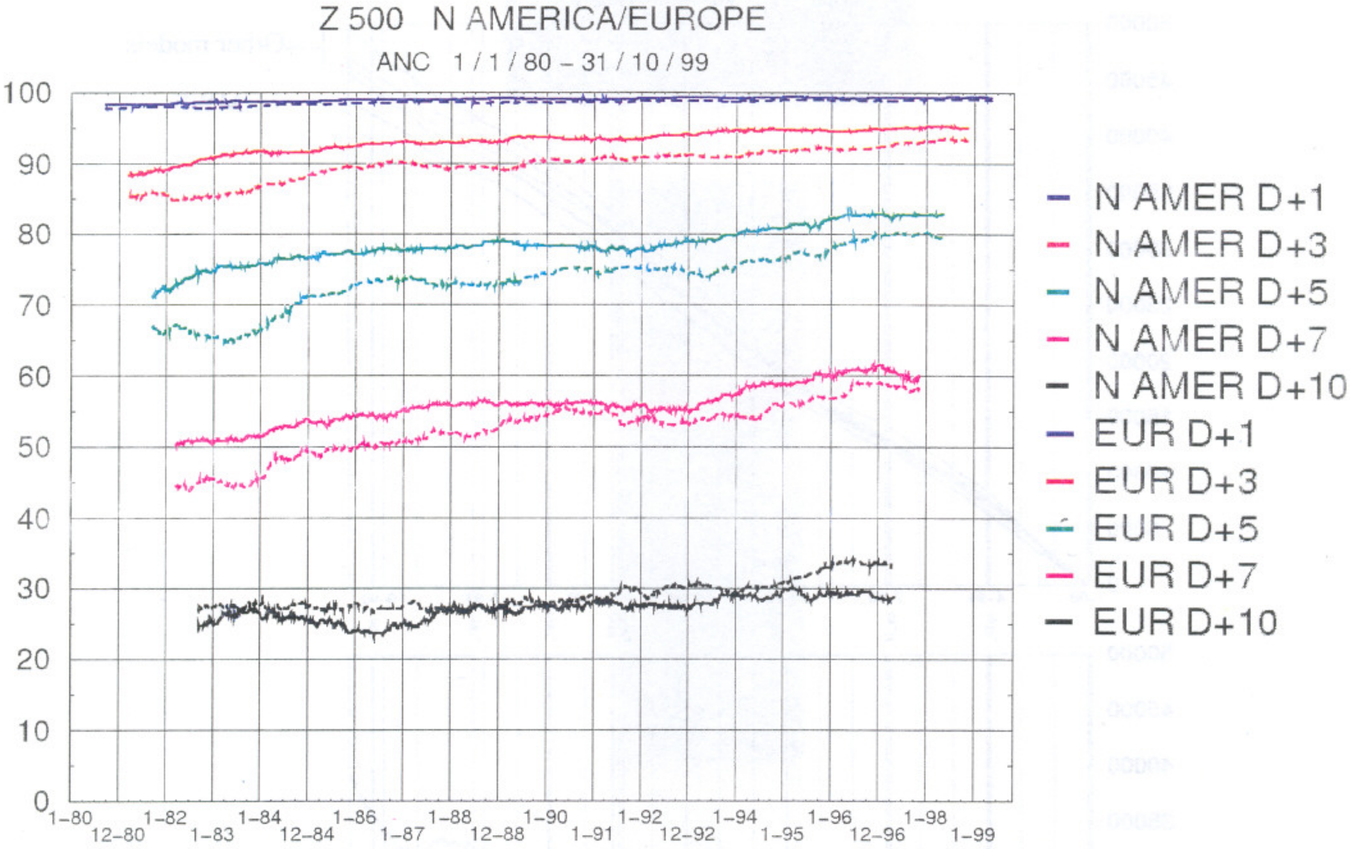


Figure 7: Time series of anomaly correlation of the 500-hPa height forecast over N. America(dashed curves) and Europe (full) over the last 20 years. Low pass median moving filters are applied with a period of 1 year (D+1, blue), 2 years (D+3, red), 3 years (D+5, green), 4 years (D+7, magenta) and 5 years (D+10, black)

3.3 Validation of Weather parameters

The verification of model precipitation, 10m wind speed, total cloud cover, 2m-temperature and specific humidity against more than 3000 SYNOP observations covering most of the world is an important component of the validation procedures routinely operated at ECMWF. The most basic information is provided in the form of bias maps and area averages over periods of time varying from one day to one month. Recently categorical scores have been introduced (frequency bias, equitable threat scores and Hansen-Kuiper (True Skill Score). When validating precipitation, however, scale representivity problems are crucial that cannot be resolved in a satisfactory way by the direct quantify these effects by making use of high resolution, climatological networks (*Ghelli and Lalaurette,* 2000).

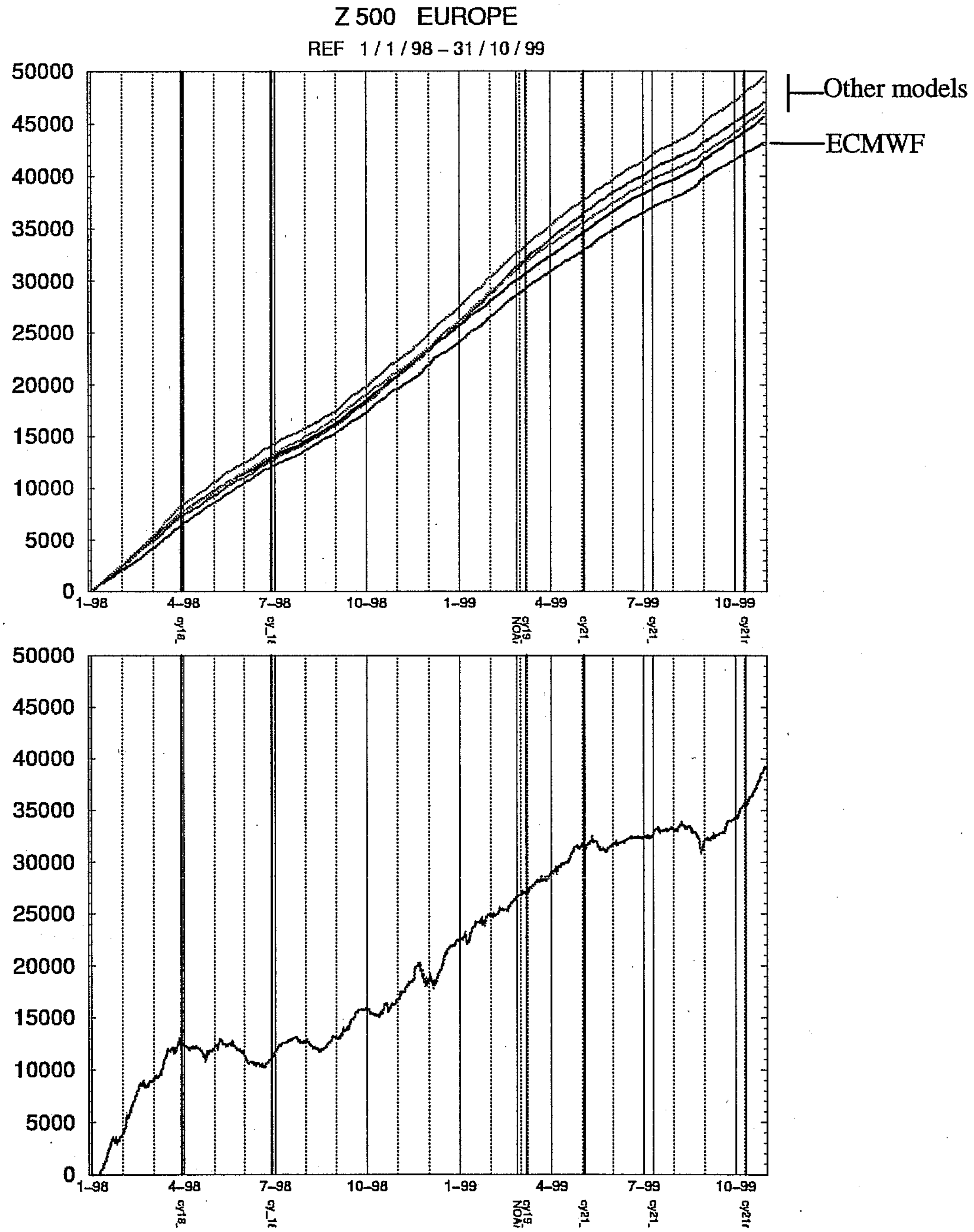


Figure 8: Upper: accumulation in time of D+5 RMSE over Europe for ECMWF and 4 other models (forecasts are retrieved from the GTS). Lower: accumulated ECMWF performance measured as the averaged accumulated error from other models minus our own. Bad spells in spring 1998, November 1998 and summer 1999 are clearly emphasised here as well as the very good winter performance. Vertical lines date recent ECMWF model changes

3.4 Validation of the EPS

It has been quite natural that in line with the early use of large scale EPS information (mainly 500-hPa height fields), the early validation procedures first targeted the ensemble mean and spread (standard deviation around the control forecast). Although expecting a high correlation between an estimator (the ensemble spread) and a stochastic quantity (the deterministic model error) is hopeless, a correlation between the spread and the model average errors when sampled according to the level of forecast spread should be found. This is indeed achieved as can be seen in Figure 9.

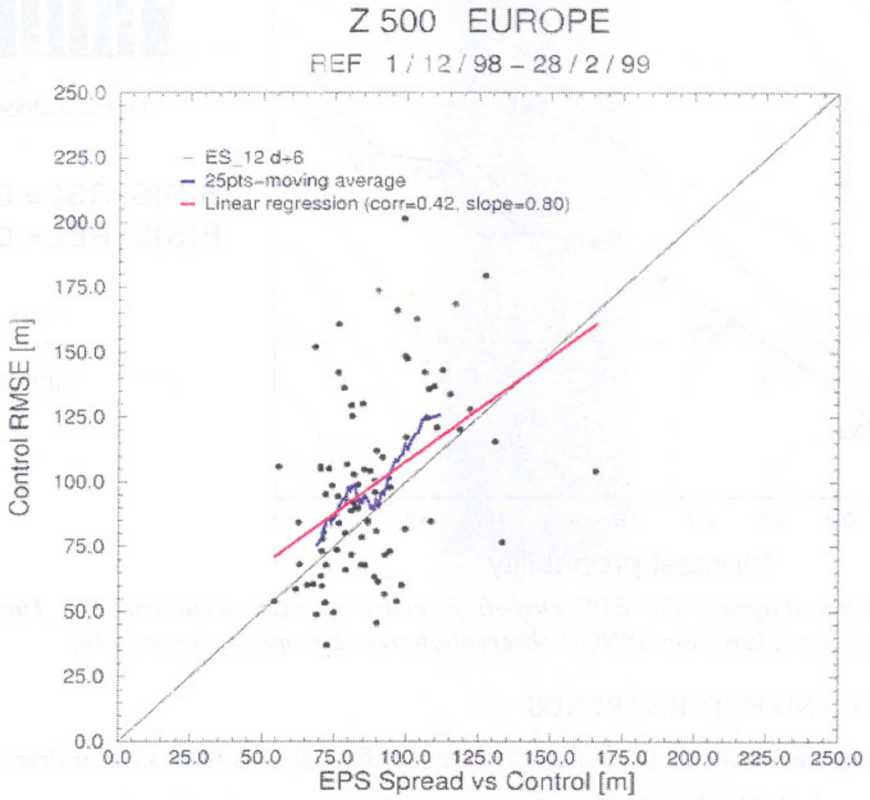


Figure 9: Spread skill relationship as exhibited from a scatter diagram accumulated over winter 1998-1999. Both the results from a linear regression and from a 25-points moving average of errors as a function of the spread are also reported.

A full validation of the probability forecast can be achieved without any Gaussian assumption for the forecast distribution by using scores such as the Brier scores which can be further partitioned into reliability and resolution components (Figure 10). While reliability is a measure of how much biased the estimation of probabilities are when compared to frequencies of occurrence of a given phenomenon, resolution is a measure of the ability of the probabilistic forecast to separate occurring from non-occurring events. Further investigations into whether the EPS detects predictable signals usually make use of Relative Operating Characteristics (ROC). Applications of these signal detection theories to the EPS are discussed later in this volume (*Richardson*, 2000)

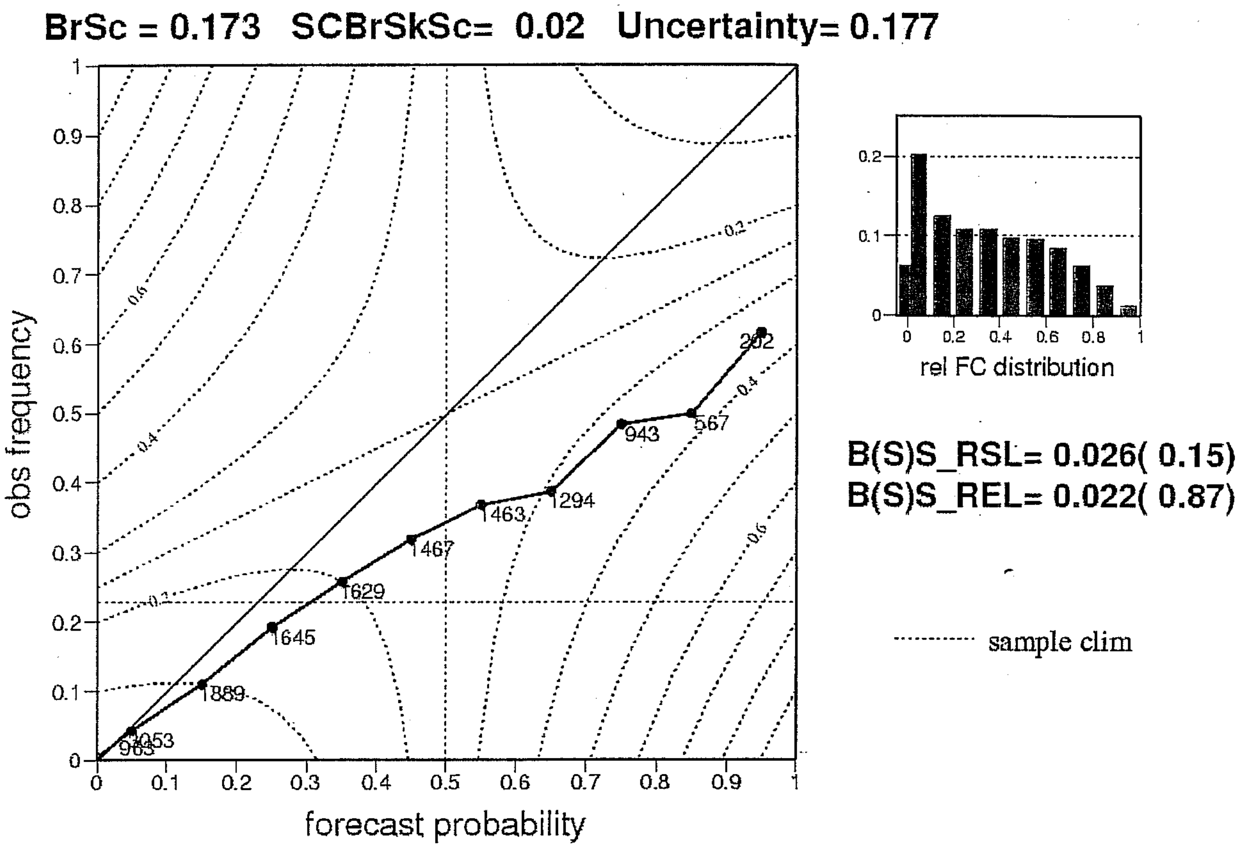


Figure 10: Reliability diagram for EPS Day+6 forecast of rain occurrence (> 1mm in 24h), verification are taken from SYNOP observation over Europe (summer 1999)

4 SUMMARY AND FUTURE TRENDS

The move from purely deterministic to stochastic numerical forecast has been a rather dramatic change in the world of Numerical Weather Prediction. It is therefore no surprise that the users needed some time before they could find the benefit they could derive from this new approach. It can now be safely said that we have left the early ages to enter a more mature one for the use of the EPS, and more generally of probabilistic products. Certainly the next millennium will be probabilistic, as new frontiers in the domain of predictability are constantly explored where a careful evaluation of the signal over noise ratio will be essential. Seasonal forecasts are an example of an area where the users will also need such a good quality, a-priory estimate of the skill.

The further we move on the way to describe explicitly the weather and to make use of the direct model output to extract value from the forecast, the more validation data we need. Certainly the international exchange of data as it has been designed in the 1950's does not address the issue of validating the model performance in predicting these parameters on a routine basis. Even for Europe, gathering reliable, real time information on precipitation fluxes in a co-ordinated way remains a goal to achieve. It has been shown however that the evaluation of model errors can be crucially dependent on the availability of such data, and the need for them will be even more crucial if reliable, early warnings for extreme events are to be expected from medium range numerical forecasts. Another challenge in that case will also be to gather the long-term climatology data that will help with the objective identification and verification of such forecasts of extreme events.

5 ACKNOWLEDGEMENTS

Anders Persson has played a key role in opening new ways of thinking medium range forecasting and interacting with forecasters from many different backgrounds and cultures. A lot of ideas in this contribution are truly his.

6 REFERENCES

Bouteloup, Y., 1997: Verification of ECMWF products in France. Report 1997 on the verification of ECMWF products in Member States and Co-operating States. 31-39

Persson, A., 2000: User's Guide of ECMWF Forecast Products. ECMWF Met.. Bulletin, *in press*

Richardson, D. S., 2000: Skill and relative economic value of the ECMWF ensemble prediction system, Quart. J. Roy. Meteor. Soc., 126, 649-668

Ghelli, A. and F. Lalaurette, 2000: Verifying precipitation forecast using upscaled observations, ECMWF Newsletter, 87, *in press*